DOCUMENT RESUME

ED 058 308 TM 001 023

AUTHOR Veldman, Donald J.
TITLE Scoring Structure and Report Generator for the 36-Item OWSC Form.
SPONS AGENCY Texas Univ., Austin. Research and Development Center for Teacher Education.
PUB DATE Jan 67
NOTE 25p.
AVAILABLE FROM R & D Center for Teacher Education, Education Annex 3.207, Univ. of Texas, Austin, Texas 78712

EDRS PRICE MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS Attitude Tests; Cluster Grouping; College Freshmen; *Computer Oriented Programs; Data Analysis; tem Analysis; *Objective Tests; *Personality Assessment; Research Methodology; *Response Style (Tests); *Scoring Formulas; Sex Differences; Social Attitudes
IDENTIFIERS *One Word Sentence Completion

ABSTRACT

This report forms the background for some current efforts to develop computer-based scoring systems for One-Word Sentence Completion data. A general discussion of response grouping problems and a three-way frequency table for data reduction are included. In addition, six "structural" response variables are defined. The report concludes with an exploratory analysis of sex differences on the 26 variables scored by the report generator, and a series of intercorrelation analyses of the internal relationships among various parts of the system. (Author/DLG)

ED058308

N-AI
TM

Scoring Structure and Report Generator for the 36-Item OWSC Form

Donald J. Veldman

Unpublished Research Memorandum, January, 1967

This report originated from a project supported by the Excellence Fund of the University of Texas and from the activities of the Computer Analysis of Personality project (NIMH 06823, 1962-1967) at the University of Texas. It forms the background for some of the current efforts to develop computer-based scoring systems for One-Word Sentence Completion (OWSC) data by the staff of the R&D Center for Teacher Education.

The report begins with a general discussion of the problem of response grouping, using a sample of the 2321 freshmen to exemplify the processes of data reduction. The final result of these procedures was a system of reducing OWSC data to a three-way frequency table: four groups of stems, six response themes, and three levels of affect/evaluation. In addition, six "structural" response variables are defined.

A report generator program is described which automatically yields a two page summary description of any OWSC protocol input to it. An example of the output of this routine is included. The report concludes with an exploratory analyses of sex differences on the 26 variables scored by the report generator, and a series of intercorrelation analyses of the internal relationships among various parts of the system.

Although never utilized in practical screening or counseling programs, the system serves as an experimental model of many of the processes to be implemented by later development projects now under way.

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY

DONALD J. VELDMAN

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

TM 001 023

SCORING STRUCTURE AND REPORT GENERATOR FOR THE 36-ITEM OWSC FORM

Donald J. Veldman

The central problem in the development of an objective scoring system for a sentence-completion instrument such as the 36-item form used in the Study of Freshmen is that of defining a reasonable procedure for reducing the variety of data to a small set of categories useful for purposes of psychological assessment. The procedure to be described in this report may be considered as an application of a four-level conceptual scheme originally described elsewhere (Veldman, 1967). This structure concerns levels at which two words (responses) may be considered the same:

(1) literal (love, love) A letter-by-letter correspondence establishes the identity of two responses.

(2) syntactic (love, loved) Grammatical variants of the same word root are considered equivalent.

(3) semantic (love, adore) Synonyms arising from different response roots are considered to have the same meaning.

(4) pragmatic (love, happy) Words are clustered at this level in terms of some particular superordinate concept, such as "affective state" for the example. The term "pragmatic" derives from the fact that the cluster is defined to serve some particular purpose of the classifier.

These four levels of classification describe rather well the sequence of procedures followed by the author in dealing with the raw data supplied by the sample of 2321 freshmen.

This report will describe one of many possible data-reduction and scoring systems which could be devised for the OWSC instrument. As noted in the discussion at the end of this paper, the reduction process

used here was perhaps carried farther than that which would be optimum for development of a flexible "score compiler" system. Still, the results of this procedure may have heuristic value for later development of a more general scoring system.

The Need for Response Grouping

The 2321 subjects responded to the 36-item sentence-completion forms with a total of 83,556 answers. Of these, 1352 (1.6%) were blanks, leaving a total of 82,204 actual response words to classify. When lists were compiled separately for each of the 36 stems, 13,743 different response words were obtained -- a reduction to 16% of the gross total. Of these words, 7971 (58%) were unique responses to the stems which elicited them, and accounted for 9.5% of the gross total. A large proportion of these "unique" responses were only grammatical variants or common synonyms for other more common responses. Even if blanks and unique responses are not considered, a total of 5772 individual words remain to be coded for use in an objective scoring system.

As described elsewhere (Peck, Menaker, and Veldman, 1966), Shirley Menaker attacked this problem in a direct fashion, although in a different context, using a 90-item form and a smaller subject sample. She assigned weights to a total of 4366 response words in order to permit automatic scoring of 25 different constructs. By assuming that words not appearing in this "dictionary" should be neutrally weighted, it was possible to avoid inclusion of an additional 2776 responses. In the scoring process which employed this weighted dictionary, all unique responses were also considered "neutral" in their implications for the 25 constructs concerned. This system has been demonstrated to be as valid and reliable as subjective scoring of the same protocols for the same constructs by individual raters.

A number of logical weaknesses are inherent in this system, however, as Dr. Menaker noted in her initial report.

First of all, the neutral weighting of unique responses probably does not bias the scores systematically. On the other hand, our experience with these data suggests that unique responses tend to be especially significant in that they often reveal unusually strong feelings or prepotent concerns on the part of the subjects. Thus, if coded, they would tend not to have neutral weights. (At the other extreme, popular responses perhaps should be coded neutral in all cases.) By leaving uniques out of the system, we are probably ignoring some powerful information.

This first problem is compounded when one considers the fact that some stems (particularly those with a few very popular responses) yield an average score considerably above the original neutral level. Thus, a person who responds with an uncommon synonym for a common response is implicitly given a below-average rating on the variable concerned.

Another consideration is the rather unreasonable amount of time and effort required of the clinician who assigns weights to the responses in the dictionary -- and ideally this should be done independently by each member of a committee and then the differences reconciled by concensus. Dr. Menaker estimates that she spent 100 hours deciding on the weighting of the 7142 responses for the 25 variables to be scored. For any new constructs, the entire list would have to be re-evaluated. For a new sentence-completion form, the entire process would have to be repeated for each new stem.

Another weakness implicit in the direct response-weight procedure appears when the system is applied to a new sample of subjects -- particularly if their verbal habits differ from those of the original sample: A somewhat larger proportion of the responses will not appear in the dictionary

-- and hence the variance of the construct scores may be severely restricted.

The procedures to be described in the following section of this report were undertaken in order to obtain a "dictionary" with greater generality than a simple compilation of responses. By deriving a final list of response-word roots, it becomes possible to pick up a substantial proportion of the unique responses. Also, the root list is only 30% as long as the corresponding list of non-unique responses which would otherwise have to be coded for scoring the sample of 2321 freshmen.

Stage One (literal equivalence)

As described in an earlier memorandum (Veldman, 1965), the 83,556 individual responses of the 2321 freshmen were compiled to form 36 lists containing a total of 5772 non-unique words. Further compilation of these 36 lists resulted in a total of 2686 response words which occurred at least once in response to one or more of the 36 stems. These words were punched into separate IBM cards to facilitate later processing.

Stage Two (syntactic equivalence)

Grammatical variants were eliminated by reducing responses to their roots, after the list of 2686 words had been alphabetically ordered. Thus, LOVE, LOVING, LOVED, LOVES, LOVER, LOVERS, LOVINGNESS, LOVABLE, AND LOVELESS were replaced by the single root LOV. The negation LOVELESS (as well as all variants with the prefixes UN-, NO-, NOT-, and NON-) is handled by a programed procedure when the list of roots is searched. The word LOVELY was not removed from the list in order that it might be separately evaluated. The particular assumptions implicit in the construction of the root list are open to argument, as well as to relatively easy adjustment if necessary. A total of 1726 roots were retained, which picked up all but 2397 (2.9%) of

the original 83,556 responses.

At this point it would have been possible to re-compile lists of occurrences of these roots for each stem, and to weight them for each of a series of constructs, yielding a scoring dictionary comparable to that developed by Dr. Menaker. For a number of reasons, we chose to continue the data reduction process through the levels of semantic and pragmatic equivalence to yield a highly abstracted response-classification system.

Stage Three (semantic equivalence)

The equating of roots carried out at this stage is even more subject to debate than was the reduction at stage two. In view of the final goal of the present process, most of these arguments would be academic in this case. However, it should be noted that construct weighting could be carried out using the synonym groups resulting from this process. Since there were roughly only 450 such groups, considerable effort on the part of the weighter(s) would be saved -- if they could agree on the synonym grouping. Each synonym group was assigned an "affective weight": 1 = negative, 2 = neutral, 3 = positive.

Stage Four (pragmatic equivalence)

This was undoubtedly the most difficult (conceptually) and time-consuming (because of false starts) of the four stages. The term "pragmatic" is apt, since the defense of any classification scheme must lie in the uses for which it is intended, and its ability to serve such purposes when tested empirically. A variety of considerations bore upon the decisions regarding this choice of six general categories or "themes":

1. Physical (body, quantity, nature, animals)
2. Roles (people, occupations, and institutions)

3. Emotional (affective states and attitudes)
4. Social (interpersonal behaviors)
5. Cognitive (ability, perception, academic)
6. Performance (success, ambition, control)

Since the final response classification system includes separate evaluation of affective tone, this aspect of the response roots was not considered in evolving the six categories. Any of these six categories could easily be subdivided into more homogeneous sets of roots, but at an unknown cost due to the decreased frequencies. The significance of this factor will become more apparent when the full design of the protocol reduction scheme is discussed in the following section of this report.

Scoring Structure for the 36-item OWSC Form

As mentioned previously, responses are to be automatically reduced to root forms and at the same time identified with regard to theme and affect level. The 36 sentence-completion stems were also grouped into four sets:

1. Self Description
2. Future Orientation and Goals
3. Stressors and Reactions
4. Social Attitudes and Interaction

The particular stems assigned to each of these categories are shown in the printout of the individual subject report which appears later in this paper. The reader should note that each of these four categories could have been further split into two subsets; this was deemed inadvisable for the same reasons that only six thematic categories were chosen for the present system.

A three-dimensional model thus serves to describe the current

scoring structure for the 36-item sentence-completion form: 4 stem sets, 6 response themes, and 3 affect levels. Theoretically, one could accumulate frequency distributions for each of the 72 cells in this model, and could then attempt to characterize a given subject or particular criterion groups in terms of their 72-element patterns. We may eventually attempt to assess the empirical validity of such a scoring system for particular external criteria, but the present report will focus upon a smaller set of summary scores derived from this 72 cell table and from the raw data of the protocol.

A set of 26 variables was defined for use in developing a report generator for individual protocols, as well as for use in comparison of criterion groups. The first six scores reflect protocol characteristics which are independent of the three-dimensional scoring procedure just described, while the remaining 20 variables are summary scores developed from this framework.

Variable 1. Number of Blank Responses. Blanks are not entered into the content-scoring framework in order to avoid any assumptions about their significance. The count, however, is reported for its value as a "tempering" variable.

Variable 2. Number of Proper Name Responses. Non-public proper names are punched "PN," and a count of these responses is made. ("John" would be coded PN, but "Kennedy" would not.) These responses are coded into the content system (Role Theme, Neutral Affect).

Variable 3. Number of Popular Responses. A table of all specific response words which were given by more than 10% of the sample of 2321 students is stored in the computer memory and referenced to obtain this score. Original response words are used in this case, as is true of all of the first six scores.

Variable 4. Number of Unique Responses. As the original responses are submitted to the root-list search system, a count is kept of the number of words which cannot be equated to any root in the dictionary. Such responses, like blanks, are not entered into the content-scoring framework for variables 7-26.

Variable 5. Number of Repeated Responses. The original list of 36 responses is examined for exact repetitions within the list. Blanks are ignored. Thus, a protocol where the word "argue" appears twice, and the word "happy" appears three times, would obtain a repetition score of three.

Variable 6. Average Response Length. Each response space in the computer is a maximum of 10 characters. The number of characters of each response is determined, summed, and divided by the number of valid responses -- since blank and proper-names are excluded in this calculation.

Affect Analysis

The next four variables are based on the affect level codes of the responses which have been retained for content analysis. (Blanks and uniques are excluded.)

Variable 7. Number of Negative Affect Responses. (code = 1) This is a simple frequency count of responses whose roots have been assigned this affect code.

Variable 8. Number of Neutral Affect Responses. (code = 2)

Variable 9. Number of Positive Affect Responses. (code = 3)

Variable 10. Average Affect Level. This score is a composite of the information represented in variables 7, 8, and 9. It is the simple average of the affect weights of all non-blank, non-unique responses.

Stem-Set Analysis

The next four variables register the average affect levels of the responses to the stems in each of the four groups of sentence-completion items of the 36-item OWSC form. Blanks and uniques are excluded from these computations.

Variable 11. Self Description Affect Level. This is an average of the affect weights for the 10 (or less) valid responses to the stems in this category.

Variable 12. Future Orientation and Goals Affect Level. This group is made up of nine stems.

Variable 13. Stressors and Reactions Affect Level. This group contains eight stems. The nature of these stems would lead to an expectation of a relatively low affect level in the typical protocol.

Variable 14. Social Attitudes and Interaction Affect Level. This group contains nine stems.

Thematic Analysis

The last 12 variables describe the frequency and the average affect level of responses whose roots had been preclassified into each of the six thematic categories. For any category, a zero frequency will automatically result in an affect level score of zero, which serves as a signal for ignoring the score in statistical analyses.

Variable 15. Frequency of Physical Theme Responses.

Variable 21. Average Affect of Physical Theme Responses. As noted earlier, bodily characteristics, animals and general quantitative descriptors are included in this category.

Variable 16. Frequency of Role Theme Responses.

Variable 22. Average Affect of Role Theme Responses. This category includes persons, occupations, and references to social institutions such as marriage.

Variable 17. Frequency of Emotional Theme Responses.

Variable 23. Average Affect of Emotional Theme Responses. Both positive and negative affective states are included in this category. The major foci are love-hate, happy-sad, and tense-relaxed.

Variable 18. Frequency of Social Theme Responses.

Variable 24. Average Affect of Social Theme Responses. These terms are descriptive of interpersonal behaviors, as opposed to emotional reactions, and reflect social morality as well as social involvement.

Variable 19. Frequency of Cognitive Theme Responses.

Variable 25. Average Affect of Cognitive Theme Responses. Perceptual traits and behaviors, intellectual ability and activity, academic references, and generalized certainty are all included within this category.

Variable 20. Frequency of Performance Theme Responses.

Variable 26. Average Affect of Performance Theme Responses. The whole range of achievement orientation, activity, and outcomes are included, along with terms concerning behavioral control and generalized evaluation.

Report Generator for Single Protocols

The scoring system described above was designed to serve as the basis for descriptive reports on individual protocols, as well as for studies of differences among criterion groups of subjects. In its present form, the system will serve the latter purpose quite adequately, but is only at an intermediate stage of development with respect to its applica-

tions to individuals. A next step would be the compilation of normative expectations for each of the 26 variables. Such tables could be studied in order to establish a network of verbal statements to be output automatically in the form of a personality description after entry of the 36 sentence completion responses from a single protocol. An intermediate version of a single-protocol report generator has been programed, and a sample of its output is shown on following pages. The first page of the report re-organizes the 36 sentence completion items into the four major stem-groups. The stems are printed (preceded by their serial numbers in the original form) with the particular subject's response words embedded. These are bracketed by asterisks. Although no such responses occurred in the example protocol, a blank would appear as *NR*, while a proper name would be printed *PN*, since that is the way they were coded when the original data cards were punched.

At the right side of this first page appear the results of the root-search procedure which is applied to each response in turn. The entry following the letter "R" is the root in the dictionary (if any) which was equated to each response. The term (UNIQUE) is printed when this search was unsuccessful -- as was the case with DISHEARTEN on stem 1. Note that negative prefixes such as UN attached to common roots are recognized as such -- as in stem 22 (third group). The negative sign is added to signify this kind of response.

The term following the letter "T" is the theme code for the selected root, which is left blank for NR or unique responses. The number following the letter "A" signifies the affect level for the particular root concerned. When a negation prefix is encountered, the original affect code is subtracted from 4 to reverse the affect level. This reversal is also

SUBJECT CODE = 0005MFT

*** PROTOCOL REDUCTION SUMMARY ***

* SELF DESCRIPTION *

1	I AM *DISHEARTEN*	R =	(UNIQUE)	T =		A =	0
17	I AM NOT *OPENMINDED*	R =	OPEN	T =	COG	A =	1
4	OTHER PEOPLE THINK THAT I AM *IRRESPONSI*	R =	IRRESPONSI	T =	PERF	A =	1
24	OTHER PEOPLE DONT KNOW THAT I AM *EMOTIONAL*	R =	EMOTION	T =	EMOT	A =	2
7	MY BEST QUALITY IS MY *INTELLIGEN*	R =	INTELLI	T =	COG	A =	3
20	MY WORST QUALITY IS MY *LAZINESS*	R =	LAZ	T =	PERF	A =	1
15	MY BODY IS *GOOD*	R =	GOOD	T =	PERF	A =	3
28	MY EMOTIONS ARE *CONTROLLAB*	R =	CONTROL	T =	PERF	A =	3
30	MY MIND IS *GOOD*	R =	GOOD	T =	PERF	A =	3
3	MY WORK HAS BEEN *EASY*	R =	EASY	T =	PERF	A =	3

* FUTURE ORIENTATION AND GOALS *

2	I SEE *HAPPINESS* WHEN I LOOK AHEAD	R =	HAPP	T =	EMOT	A =	3
19	TEN YEARS FROM NOW I WILL BE *MARRIED*	R =	MARR	T =	ROLE	A =	2
23	I AM *CERTAIN* ABOUT WHO I AM AND WHERE I AM GOING	R =	CERTAIN	T =	COG	A =	3
36	REAL-LIFE STORIES USUALLY HAVE *REAL* ENDINGS	R =	REAL	T =	PHYS	A =	3
16	THE MOST IMPORTANT THING IN LIFE IS *HAPPINESS*	R =	HAPP	T =	EMOT	A =	3
26	WHAT I WANT MOST IS *HAPPINESS*	R =	HAPP	T =	EMOT	A =	3
9	*WORK* LEADS TO SUCCESS	R =	WORK	T =	PERF	A =	3
29	*WORK* LEADS TO SUCCESS	R =	WORK	T =	PERF	A =	3
12	SECRETLY, I OFTEN DREAM ABOUT *ANXIETIES*	R =	ANXI	T =	EMOT	A =	1

* STRESSORS AND REACTIONS *

6	I DONT LIKE TO THINK ABOUT *STUDYING*	R =	STUDY	T =	COG	A =	2
10	*HOMOSEXUAL* MAKES ME ANGRY	R =	(UNIQUE)	T =		A =	0
14	I AM AFRAID OF *FAILURE*	R =	FAIL	T =	PERF	A =	1
32	*HOME* MAKES ME FEEL SAD	R =	HOME	T =	ROLE	A =	2
8	DARKNESS IS *GOOD*	R =	GOOD	T =	PERF	A =	3
22	WHEN AN ANIMAL IS WILD, IT IS *UNCONTROLL*	R =	-CONTROL	T =	PERF	A =	1
35	I *IMPROVE* WHEN PUT UNDER PRESSURE	R =	IMPROV	T =	PERF	A =	3
33	I *DRINK* WHEN I LET GO	R =	DRINK	T =	PHYS	A =	2

* SOCIAL ATTITUDES AND INTERACTION *

5	MEN OFTEN *DESIRE* WOMEN	R =	DESIR	T =	PERF	A =	2
18	WOMEN OFTEN *SATISFY* MEN	R =	SATISF	T =	EMOT	A =	3
21	I HAVE *MANY* FRIENDS	R =	MANY	T =	PHYS	A =	2
34	OTHER PEOPLE MAKE ME *EXAGGERATE*	R =	(UNIQUE)	T =		A =	0
11	THE AVERAGE PERSON IS *SUCCESSFUL*	R =	SUCCE	T =	PERF	A =	3
25	MOST MEN ARE *STRONG*	R =	STRONG	T =	PHYS	A =	3
13	MOST WOMEN ARE *BITCHES*	R =	BITCH	T =	SOC	A =	1
31	MY MOTHER IS *SWEET*	R =	SWEET	T =	SOC	A =	3
27	MY FATHER IS *GOOD*	R =	GOOD	T =	PERF	A =	3

SUBJECT CODE = 0005MFT

* PROTOCOL CHARACTERISTICS *

0. BLANK RESPONSES.

0. PROPER-NAME RESPONSES.

9.00 POPULAR RESPONSES.

3.00 UNIQUE RESPONSES.

6.00 REPEATED RESPONSES.

7.08 = AVERAGE RESPONSE LENGTH.

* AFFECT ANALYSIS *

7.00 NEGATIVE RESPONSES.

7.00 NEUTRAL RESPONSES.

19.00 POSITIVE RESPONSES.

2.36 = AVERAGE AFFECT LEVEL.

* STEM-SET ANALYSIS *

SELF DESCRIPTION	A = 2.22
FUTURE ORIENTATION AND GOALS	A = 2.67
STRESSORS AND REACTIONS	A = 2.00
SOCIAL ATTITUDES AND INTERACTION	A = 2.50

* THEMATIC ANALYSIS *

PHYSICAL	F = 4.00	A = 2.50
ROLES	F = 2.00	A = 2.00
EMOTIONAL	F = 4.00	A = 2.50
SOCIAL	F = 2.00	A = 2.00
COGNITIVE	F = 4.00	A = 2.25
PERFORMANCE	F = 15.00	A = 2.40

always applied to the response to stem 17 (first stem group). The theme and affect codes form the primary data for construction of most of the 26 scores on the second page of the report. An accidental, but interesting, artifact of the 36-item form appears in group two with stems 9 and 29, which are identical. They do not always elicit identical responses, however.

The second page of the report contains the 26 summary scores listed earlier, with appropriate labeling. Although one could employ these scores in studies comparing criterion groups such as males and females -- as we will see later in this paper -- direct interpretation of the scores for a single subject would be dangerous at best, without some sort of normative data. This particular subject seems to have used a large number of "performance" terms, for instance, but it is impossible for us to say at this point how unusual such a score might be. Even with normative data available for each separate variable, we will probably design our later report-writing routines so as to qualify statements based on one score in situations where other scores (such as the blank count) are unusual. Thus, even after obtaining normative tables for each of these 26 variables, a great deal of further work will be necessary to produce a sufficiently sophisticated report-generating program.

Male-Female Differences

The 26 basic scores described earlier were developed for each of the 2321 protocols of the freshman sample, and the 1362 males were compared to the 959 females on each of these variables. Table 1 contains the results of the analyses of variance. The reduced N for some variables was necessitated by the zero affect-level scores which resulted when a subject gave no responses in the particular category being considered.

These results clearly indicate that females use more popular response words, and fewer unique response words, than do males. Females also repeat response words less often within their protocols, and tend to use longer words as responses. Female subjects as a group use more negative-affect responses than do males. There is a tendency for males to use more neutral responses, but there is no sex difference in the use of positive-affect words. These results are in essential agreement with those reported in an earlier memorandum (Veldman, 1965).

The breakdown of data in the present system permits us to explore further the matter of sex differences in response to various types of stems. It is quite clear from Table 1 that the females' use of negative response words occurs most heavily on those stems which present stressors or stress reactions. There is also a clear tendency for the female sample to use more negative words for self-description, although this difference is not as strong.

When the response words are classified according to the six-theme system, significant sex differences in frequency of usage occur for every one of the six categories. Males use the Physical, Cognitive, and Performance themes more often, while females use the Role, Emotional, and Social themes more frequently.

The last six variables register the average affect levels of the words in each of the six thematic categories. Subjects who used no words in a particular category were excluded from the corresponding analysis of variance. No significant differences appeared for the Social or Performance themes. Males responded with more positive affect in the Physical, Emotional, and Cognitive thematic categories, while females gave more positive Role theme responses.

Although many of the assignments of response words to thematic categories are open to question, it is quite clear that males and females respond to the sentence completion instrument in distinctively different ways. Beyond the strong suggestion that females respond with generally greater verbal facility to this task, patterns of response themes appear which are clearly sex-role appropriate. The fact that females use more words with negative affective loadings is not easily explained, however, and is contrary to what might have been predicted from the stereotypic college coed personality pattern. That these negative words appear most frequently in response to stems presenting stressors and stress reactions only partially explains the phenomenon, since the same sex difference occurs with self-descriptive items.

These strong sex differences in response tendencies dictate the development of separate normative tables for males and females for use with future report-writing procedures.

Table 1. Males vs. Females Analyses of Variance

	Variable	Ns	Males	Females	F	P
1.	Number of Blanks	(1362, 959)	.64	.50	1.74	.1842
2.	Number of Proper Names	(1362, 959)	.11	.12	.14	.7084
3.	Number of Populars	(1362, 959)	6.48	7.83	113.63	<.00005
4.	Number of Uniques	(1362, 959)	1.14	.88	19.30	.0001
5.	Number of Repetitions	(1362, 959)	4.17	3.65	30.08	<.00005
6.	Average Response Length	(1362, 959)	6.70	6.78	9.43	.0026
7.	Number of Negative Affect	(1362, 959)	10.64	11.36	23.95	<.00005
8.	Number of Neutral Affect	(1362, 959)	7.55	7.37	2.35	.1210
9.	Number of Positive Affect	(1362, 959)	16.03	15.89	.72	.5974
10.	Average Affect Level	(1362, 959)	2.16	2.13	11.08	.0013
11.	Self-Description Affect	(1362, 958)	2.16	2.12	8.67	.0036
12.	Future-Goals Affect	(1362, 959)	2.50	2.49	1.24	.2644
13.	Stress Reactions Affect	(1362, 959)	1.61	1.54	30.36	<.00005
14.	Social Attitudes Affect	(1362, 959)	2.28	2.30	1.05	.3073
15.	Physical Theme Count	(1362, 959)	5.42	5.08	12.88	.0006
16.	Role Theme Count	(1362, 959)	2.69	2.92	9.76	.0022
17.	Emotional Theme Count	(1362, 959)	7.31	9.21	239.99	<.00005
18.	Social Theme Count	(1362, 959)	4.13	4.72	40.22	<.00005
19.	Cognitive Theme Count	(1362, 959)	4.30	3.98	13.21	.0006
20.	Performance Theme Count	(1362, 959)	10.38	8.71	159.53	<.00005
21.	Physical Theme Affect	(1359, 953)	1.96	1.92	10.04	.0020
22.	Role Theme Affect	(1242, 916)	1.88	1.95	52.32	<.00005
23.	Emotional Theme Affect	(1360, 958)	2.35	2.29	14.81	.0003
24.	Social Theme Affect	(1329, 949)	1.97	1.99	.38	.5437
25.	Cognitive Theme Affect	(1348, 943)	2.15	2.07	12.07	.0009
26.	Performance Theme Affect	(1362, 958)	2.27	2.27	.14	.7107

Intercorrelation of Scores

Separate 26 x 26 correlation matrices were computed for the male and female subsamples. They will not be reproduced here, since most of the coefficients reflect only the nature of the response categorization system. Blank and unique responses were excluded from this system, however, and the relationships between these counts and the various affect-score averages are shown below:

	Blank Count		Unique Count	
	Males	Females	Males	Females
All Responses	.01	.01	-.12	-.09
Self Stems	-.02	.03	-.05	-.03
Future Stems	.01	.04	-.15	-.13
Stress Stems	.05	.00	.11	.14
Social Stems	-.02	-.05	-.18	-.17
Physical Responses	.02	.01	-.02	.01
Role Responses	-.05	-.01	.01	-.03
Emotional Responses	.01	-.04	-.09	-.05
Social Responses	.03	.10	-.04	-.04
Cognitive Responses	-.03	.08	.04	-.06
Performance Responses	.00	-.07	-.08	.01

It is quite clear from these data that the giving of blank responses is not associated with the affect level of responses to other stems in general, nor within any of the stem groups. There appears to be a slight tendency for females who give many blanks to also give positive social and cognitive responses, but negative performance responses.

The tendency to use unique response words seems to be associated slightly with the use of negative-affect words on other stems generally. This trend is especially true for stems concerning future orientation and social interaction, but the reverse appears with stems concerning stress stimuli. A high count of uniques tends slightly to be associated with lower emotional and performance-response affect with males, but no relationships appear for the female sample.

The intercorrelations of the affect averages for the four stem groups are shown in the next table. For both sexes, the affect level for self-description is associated with that for future orientation and for social interaction, but is not related to the affect level for stress-stem responses. The social interaction and future orientation affect levels are also associated to about the same degree. The lack of relationship between stress-stem responses and the others is probably due to the fact that these stems are constructed to elicit negative reactions.

Intercorrelation of Affect Averages for Stem Groups

	Self	Future	Stress	Social	
Self	--	.29	.01	.24	Males above diagonal
Future	.29	--	-.13	.21	Females below diagonal
Stress	.00	-.03	--	-.07	
Social	.18	.22	-.13	--	

The next table of intercorrelations reflects the associations among the affect levels of the various response categories. In general, these coefficients are surprisingly low. The strongest relationships are between emotional responses on the one hand, and cognitive and performance responses

on the other. These protocols certainly do not seem to be homogeneous in mood. To some extent these coefficients suggest that the present scoring system is not sensitive enough to the nature of the particular stems eliciting the response words. For example,

"Stupidity makes me angry."

and "My best quality is my intelligence."

yield opposite affect scores in the cognitive response category, yet they certainly are consistent with each other. Such internal contradictions would reduce any cross-category relationships.

Intercorrelation of Affect Averages for Response Categories

	Physical	Role	Emotional	Social	Cognitive	Performance
Physical	--	-.02	.07	.09	.05	.01
Role	.00	--	-.02	.00	-.05	-.03
Emotional	.09	-.06	--	.10	.11	.13
Social	.01	.01	.08	--	.08	.06
Cognitive	.07	-.03	.16	.05	--	.13
Performance	.03	-.03	.14	-.01	.09	--

Discussion

The system described in this progress report has a number of weaknesses which are known at this time; others will undoubtedly become apparent as work continues to improve the flexibility and precision of the scoring procedures. It may be worthwhile to specify some of these directions for efforts to improve the present system.

One weakness is the relatively large number of roots retained which contain common prefixes not cut out by the present ROOT routine.

An earlier memorandum described a more complex root-cutting procedure (Veldman, 1965), and we will probably modify the present system in that direction in order to reduce the size of the root dictionary and increase its range of applicability. We also may wish to consider the practicality of hand-screening future data to reduce multi-word responses to key words prior to punching.

The usefulness of a system like the one described here, which reduces the responses to only six thematic categories and the stems to four groups, is open to serious question as a general approach to the goal of developing a scoring system for the OWSC data. The primary question is one of the adaptability of the system to the variety of purposes to which sentence completion data may be applied in various research situations. It has been suggested that the present system is too concise to allow users to get meaningful answers to particular questions they might hope to answer with the instrument. Suppose one wished to know the subject's attitude toward his parents. About the closest score to this area would be the Role Affect level score -- which, of course, registers attitudes toward a much wider variety of role-objects and institutions. Lack of ability to answer specific questions within the system framework is not the only weakness; one might also wish to develop broadly-based scores of a different nature than those which are yielded by the current system -- even in its 72-cell form.

One way of building a more flexible system will be outlined here. We intend to implement this approach with the Freshmen Study data in order to test its feasibility. The system will operate as outlined in the present report up to the semantic stage of response equivalencing -- with perhaps an increase in the prefix-chopping power of the ROOT routine.

Instead of the rather extreme reduction to six themes at the pragmatic level, however, we will employ somewhat fewer than 100 thematic categories of synonymic root-sets. Each synonymic set will also be coded for "dimensionality." The notion here is that some thematic categories can be organized along an affect dimension. For instance, "love-hate" might form a thematic category, within which LOVE and ADORE might be coded "5" for dimensionality, while LIKE would be coded "4", HATE would be coded "1", etc. Other thematic categories would also contain terms coded for "dimensionality;" but the dimension would not be affective in nature. For instance, in the thematic category "parents," FATHER would be coded "4", and MOTHER would be coded "2"; the numerical codes would only signify the dichotomous nature of the theme. Ambiguous terms like PARENT would be coded "3". Still other themes would be dimensionalized with a more general evaluative loading, such as "size."

Use of such a less-than-100-theme system, with further differentiation by a one-digit dimensionality code, would permit the score designer considerable versatility. When we then add to the system a two-digit code for the particular sentence-completion stem concerned, each protocol can be described in terms of 36 five-digit codes. Scores can be defined as sums of dimensionality codes for particular thematic categories and/or particular stems, or in any of a wide variety of other ways.

To make such a data-reduction system most useful to psychologists, it would be necessary to develop a computer program which would accept formulas for specific scores, written in a concise notation. Some "standard" formulas could be developed initially as a general-purpose scoring system, but the organization of the data-reduction scheme and the program for translating and implementing score formulas would enable rapid and

economical scoring of any set of protocols for an almost infinite variety of special purposes.

Finally, we may consider the advisability of designing a new OWSC form to recommend in all future research situations. Accumulated experience with the data provided both by the 90 and 36-item versions strongly suggests that it would be best to limit the stem formats to two more-or-less standard arrangements. The most straightforward type of stem presents an object and asks for a descriptor. For instance, "My father is _____." The reverse arrangement has also been found useful: "I am proud of my _____." Stem formats other than these within a protocol complicate the root-reduction system, and seldom contribute data which cannot be obtained from the formats shown above. To increase the usefulness of the OWSC as a screening instrument, we should probably construct stems which would be useful in a wide variety of research and individual-assessment situations, with high school, college and adult subjects, and yet short enough (perhaps 20 - 30 stems) to be suitable for use in mass testing programs.

REFERENCES

Veldman, D. J. FORTRAN Programming for the Behavioral Sciences. Holt, Rinehart, and Winston, 1967.

Peck, R. F.; Menaker, Shirley; and Veldman, D. J. Computer analysis of sentence completion data. American Psychologist, 1966, 21, 678.

Veldman, D. J. A study of freshmen: introduction. Computer Analysis of Personality Project, August, 1965 (mimeo).

Veldman, D. J. Response reduction to word roots. Computer Analysis of Personality Project, October, 1965 (mimeo).